

▼ Архитектура нейронной сети RetinaNet

Архитектура свёрточной нейронной сети (СНС) RetinaNet состоит из 4 основных частей, каждая из которых имеет своё назначение:

- **Backbone** – основная (базовая) сеть, служащая для извлечения признаков из поступающего на вход изображения. Данная часть сети является вариативной и в её основу могут входить классификационные нейросети, такие как ResNet, VGG, EfficientNet и другие;
- **Feature Pyramid Net (FPN)** – свёрточная нейронная сеть, построенная в виде пирамиды, служащая для объединения достоинств карт признаков нижних и верхних уровней сети, первые имеют высокое разрешение, но низкую семантическую, обобщающую способность; вторые – наоборот;
- **Classification Subnet** – подсеть, извлекающая из FPN информацию о классах объектов, решая задачу классификации;
- **Regression Subnet** – подсеть, извлекающая из FPN информацию о координатах объектов на изображении, решая задачу регрессии.

▼ Backbone

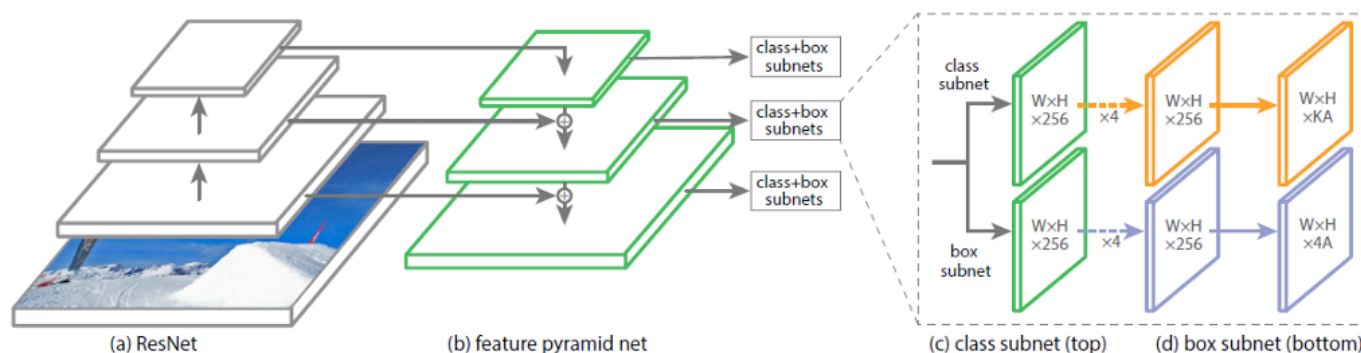


Рисунок 1 – Архитектура RetinaNet с backbone-сетью ResNet

▼ Feature Pyramid Network

Feature Pyramid Network состоит из трёх основных частей: восходящий путь (bottom-up pathway), нисходящий путь (top-down pathway) и боковые соединения (lateral connections).

последовательность свёрточных слоёв с уменьшающейся размерностью, в нашем случае – backbone сеть.

Верхние слои свёрточной сети имеют большее семантическое значение, но меньшее разрешение, а нижние наоборот (рис. 2). Bottom-up pathway имеет уязвимость при извлечении признаков – потеря важной информации об объекте, например из-за зашумления небольшого, но значимого, объекта фоном, так как к концу сети информация сильно сжата и обобщена.

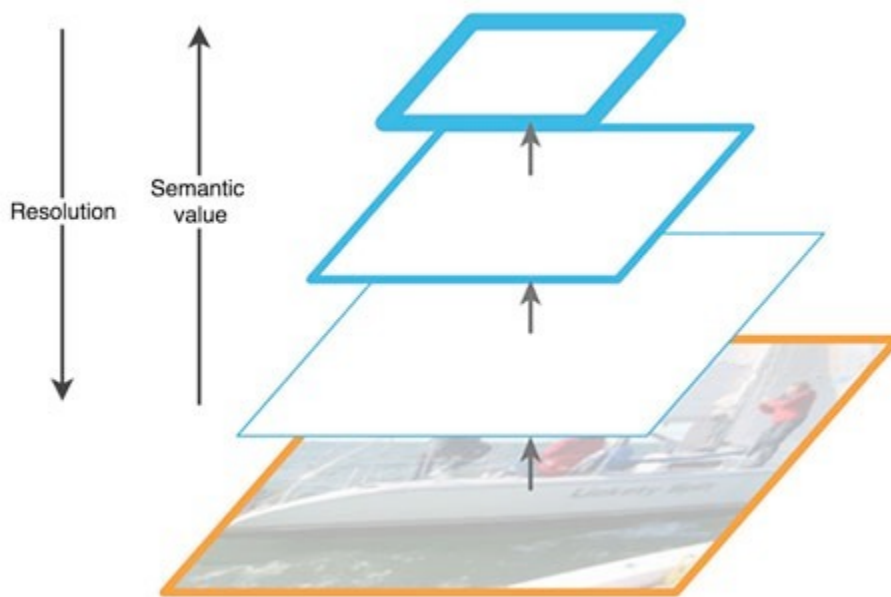


Рисунок 2 – Особенности карт признаков на разных уровнях нейросети

Нисходящий путь также представляет собой «пирамиду». Карты признаков верхнего слоя этой пирамиды имеют размер карт признаков верхнего слоя bottom-up пирамиды и увеличиваются вдвое методом ближайшего соседа (рис. 3) по направлению вниз.

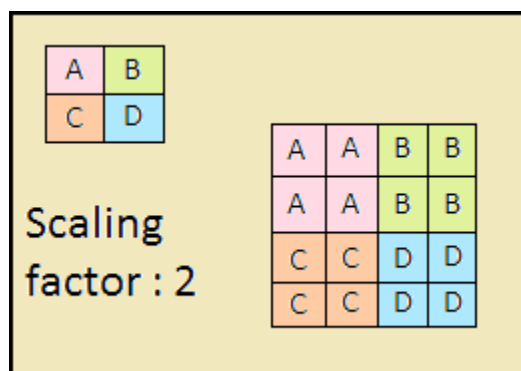


Рисунок 3 – Увеличение разрешения изображения методом ближайшего соседа

Таким образом в top-down сети каждая карта признаков вышележащего слоя увеличивается до размеров карты нижележащего. Помимо этого, в FPN присутствуют боковые соединения, это означает, что карты признаков соответствующих слоёв bottom-up и top-down пирамид поэлементно складываются, причём карты из bottom-up проходят свёртку 1×1 .

Боковые соединения решают проблему затухания важных сигналов в процессе прохода по слоям, совмещая семантически важную информацию, полученную к концу первой пирамиды и более детальную информацию, полученную в ней ранее.

Далее, каждый из полученных слоёв в top-down пирамиде обрабатывается двумя подсетями.

Подсети классификации и регрессии

Третьей частью архитектуры RetinaNet являются две подсети: классификационная и регрессионная. Каждая из этих подсетей образует на выходе ответ о классе объекта и его расположении на изображении. Рассмотрим принцип работы каждой из них.

Разница в принципах работы рассматриваемых блоков (подсетей) не отличается до последнего слоя. Каждый из них состоит из 4 слоёв свёрточных сетей. В слое формируются 256 карт признаков. На пятом слое количество карт признаков изменяется: регрессионная подсеть имеет $4A$ карт признаков, классификационная – KA карт признаков, где A – количество якорных рамок (подробное описание якорных рамок в следующем подразделе), K – количество классов объектов.

В последнем, шестом, слое каждая карта признаков преобразуется в набор векторов. Регрессионная модель на выходе имеет для каждой якорной рамки вектор из 4 значений, указывающих смещение целевой рамки (англ. ground-truth box) относительно якорной. Классификационная модель имеет на выходе для каждой якорной рамки one-hot вектор длиной K , в котором индекс со значением 1 соответствует номеру класса, который нейросеть присвоила объекту.

Сравнение RetinaNet и YOLOv3

Модель			Функция потерь Loss	Точность предсказания mAP (%)	Среднее время обработки изображения
Detector	Backbone	Anchors			
RetinaNet	ResNet-34	9	0.595	90.4%	89 ms
	ResNeXt-101	9	0.401	92.8%	121 ms
YOLOv3	Darknet-53	6	9.311	42.6%	76 ms
		9	7.360	59.8%	76 ms

Архитектура RetinaNet показала конкурентоспособные результаты по отношению к YOLOv3. Самая «легкая» модель RetinaNet превзошла результаты тестирования YOLOv3 с большим преимуществом. RetinaNet с архитектурой ResNet-34 всего на 7% медленнее, чем YOLOv3, но превосходит лучшее значение mAP сети YOLOv3 на 30,6 процентов.

Источники:

1. Архитектура нейронной сети RetinaNet [Ссылка](#)
2. Научная статья. авт. Болховитина Е.И. [Ссылка](#)

[Платные продукты Colab](#) - [Отменить подписку](#)